On the Sampling Frequency of Human Mobility

Panagiota Katsikouli, Aline Carneiro Viana, Marco Fiore, Alberto Tarable

On the Sampling Frequency of Human Mobility

Panagiota Katsikouli*, Aline Carneiro Viana†, Marco Fiore‡, Alberto Tarable‡
* University of Edinburgh, † Inria, ‡ CNR-IEIIT
Email: *p.katsikouli@sms.ed.ac.uk, †Aline.Viana@inria.fr, ‡marco.fiore@ieiit.cnr.it, ‡alberto.tarable@ieiit.cnr.it

***Abstract*—In this paper, we aim at answering the question *"at what frequency should one sample individual human movements so that they can be reconstructed from the collected samples with minimum loss of information?"*. Our quest for a response unveils *(i)* seemingly universal spectral properties of human mobility, and *(ii)* a linear scaling law of the localization error with respect to the sampling interval. We conduct analyses using fine-grained GPS trajectories of 119 users worldwide. Our findings have potential applications in ubiquitous computing and mobile service design, in terms of energy efficiency, location-based service operations, active probing of subscribers' positions in mobile networks and trajectory data compression.**

I. Introduction

Over the past few years, the pervasive usage of smart devices and location-tracking systems has made it possible to study and understand human mobility at unprecedented scales. An important feature that was found to characterize human mobility is regularity; we tend to follow the same patterns over and over, and we do so in ways that are clearly periodic [1]. Regularity is easily found in the movements of most individuals: as an example, consider Fig. 1, which shows heatmaps of the locations visited by three random users in the dataset employed in our study. Although these plots convey three weeks of data, a small set of frequently visited places emerges for all users, along with systematic paths connecting them. Likewise, Fig. 2 illustrate the temporal dimension of regularity for the same users: a clear periodicity emerges from the time series of the visited locations.

In this paper we investigate whether the regularity of human mobility entails the possibility of sampling individual movements at reduced constant frequencies, while allowing for the reconstruction of trajectories that retain a vast portion –if not all– of their original level of detail. Intuitively, periodic visits to a limited set of important places through repeated routes may be captured with a smaller sampling effort than, e.g., a completely random mobility. Identifying suitable frequencies for human mobility sampling would have applications in a number of fields, including but not limited to:
(i) mobile computing, where overly frequent GPS localization unnecessarily reduces the battery life of mobile devices;
(ii) location-based service design, where unwarranted users' position data collection raises significant privacy concerns;
(iii) cellular networks, where active probing of subscribers' positions is a costly task whose rate must be duly optimized;
(iv) trajectory data compression, where information loss must be minimized.

Overall, our problem is equivalent to posing the question *"at what frequency should one periodically sample individual human movements so that they can be reconstructed from the collected samples with minimum loss of information?"*.

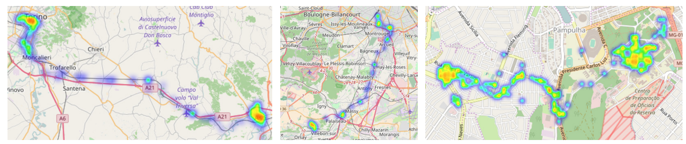


Fig. 1: Heatmaps of locations visited by three distinct users during three weeks: humans tend to commute between a limited set of specific locations. Figure best seen in color.

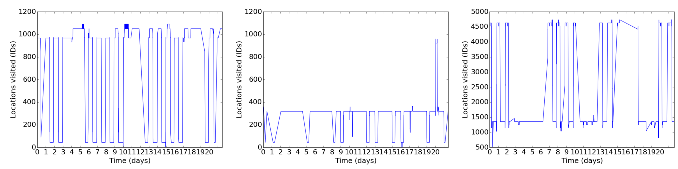


Fig. 2: Location time series for three distinct users during three weeks: humans tend to revisit locations in a periodic fashion. The visited locations are mapped to sequential identifiers, upon discretization on a regular grid of 50 meters step.

To respond, we adopt a signal processing approach, by considering mobility patterns as signals over time, and carrying out a spectral analysis of human mobility. We find that the spectra of the movements of 119 individuals have very similar, flat shapes that suggest the absence of convenient sampling frequency thresholds – even specific to single users – beyond which the error in the reconstructed trajectories drops significantly.

Stimulated by this finding, we carry out a quantitative analysis of the user localization error in movements reconstructed from regular sampling at different periodicities. Our results unveil a linear scaling law of the error with respect to the span of the constant sampling interval. This law corroborates the outcome of the spectral analysis and has significant practical implications, as it controls the trade-off between accuracy and cost of measurements of human mobility.

II. Related work

Spatial data trajectory compression is a widely addressed research subject, where the objective is to maintain the trajectory shape (see [2], [3] and references therein). We aim at preserving the temporal dimension of movements as well. Consider the toy example in Fig. 3, where a user leaves home, trains at the gym before work and, later goes to a take-away restaurant. The shape of the spatial trajectory could be well approximated as the sequence of home, junctions B and C, and take-away locations: map-matching based on these cardinal points would allow the description of the whole movement. However, individuals visit places for a purpose, carrying out activities that have different durations. In our example, the size of the circle around each location is proportional to the amount of time spent there. Our purpose is to recreate the complete original mobility of the user, including these temporal features.

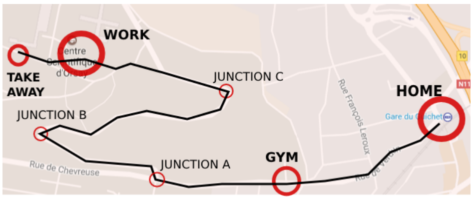


Fig. 3: Toy example of a mobility trajectory. Labels denote important locations or turning points. The size of the circle around each location is proportional to the amount of time spent there by the user.

Our problem is also different from sampling to detect important locations [4], [5], or from simplifying GPS trajectories to preserve semantics of locations [6], [7]. In the example of Fig. 3, important location detection is solved by sampling the trajectory so as to model the original distribution of time spent at home, work, gym, and take-away. However, approaches for the detection of such frequently visited places ignore the time ordering of visits, and do not capture transitions between frequent locations. Instead, our holistic perspective accounts for all these characteristics.

Finally, we are not interested in maintaining locations that would impose a great change in the original direction of the trajectory, if absent [8]; nor we address the similar problem of calculating the current position of a target based on its travelled distance and direction of movement, known as dead reckoning [9], [10]. Indeed, we are not interested in simplifying a pre-recorded GPS trajectory but to find convenient sampling frequencies for human trajectory data.

To the best of our knowledge, this is the first work to thoroughly study the problem of finding a good constant sampling frequency at which to sample human mobility so that users' complete movements can be accurately reconstructed.

III. Reference dataset

Our study employs a dataset of real-world individual mobility data extracted from three different sources.

- MACACO data was collected between July 2014 and December 2016 as part of the European collaborative project MACACO, funded by the EU CHIST-ERA program[1]. A dedicated application, running on smartphones of volunteers in several countries in Europe and South America, recorded GPS positioning information at regular time intervals, typically from one to five minutes. Due to data privacy regulations in France, where the GPS logs are hosted, the data is not publicly available.
- OpenStreetMap (OSM) data was collected by volunteers who recorded and uploaded their trajectories as a contribution to the OSM database[2]. The OSM project is a global crowd-sourcing initiative, aiming at mapping the whole world surface thanks to the activity of a vast user community. GPS traces uploaded by participants typically feature 1-Hz frequency. They are freely available on the official OSM project website.
- Geolife data was collected in Beijing by Microsoft Research Asia from April 2007 to August 2012 [11]. It consists of GPS trajectories recorded through different GPS loggers and smartphone apps. Although sample rates vary significantly across users and time periods, the vast majority of GeoLife positioning data is recorded at intervals from one to five seconds. GeoLife traces are publicly available on the official project website.

[1]https://macaco.inria.fr/macacoapp/

[2]https://www.openstreetmap.org

Dataset	Users	Weeks
MACACO	19	164
OpenStreetMap	4	7
GeoLife	96	881

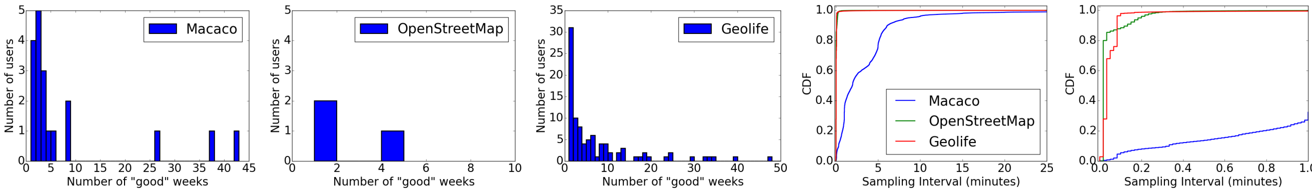
TABLE I: Per-source users and weeks in the reference dataset.

The GPS trajectories in our dataset cover sensibly different geographical and temporal spans, even for data coming from the same source. Depending on the user, movements can encompass a single city or multiple continents, over time intervals ranging from days to years. Moreover, the quality of the data for a single user is typically very heterogeneous over time, with periods of days or weeks where GPS logs are erroneous or completely absent. In order to build a consistent reference dataset, we segmented the mobility traces of all users into one-week trajectories[3], and analysed them separately. During each week, we bounded the mobility of each individual to the regions where the activity is concentrated. One-week trajectories and bounded regions avoid biases introduced by singularities, such as international journeys performed once in months. Clearly, bounded regions can create temporal gaps in the weekly traces, whenever users visit areas outside them; however, this effect is marginal, as we found that users spend 85% to 100% of their time within their weekly bounding region. Depending on the user, bounding regions span from 400 to $3,000\ km^2$.

We then filtered the one-week geographically-bounded trajectories based on their quality and retained only the trajectories that contain complete GPS records in at least six out of seven distinct week days. Ultimately, our reference dataset is composed of 1,052 weeks of mobility of 119 different individuals. Tab. I provides a break down of these numbers on a per-source basis. A legitimate question is whether the data is dominated by a few users, i.e., if the majority of weeks refers in fact to a limited set of users, which could bias the analysis. Fig. 4(a)-(c), which portray the distribution of the number of weekly trajectories, show that this is not the case. Indeed, the vast majority of users contribute one to ten weeks of movement data, and the few users who exceed that range provide around 50 weeks of mobility at most. Overall, our reference dataset encompasses a quite diverse base of individuals.

The different techniques employed to collect the GPS positioning information lead to uneven recording intervals across, and even within, the original data sources. In addition to this, weekly trajectories have temporal gaps due to offline GPS receivers, interruptions in the data collection service, or users travelling outside the bounding regions we introduce. Fig. 4(d) shows the cumulative distribution function (CDF) of the sampling intervals observed in all one-week trajectories of our reference dataset. We remark that in almost all cases such intervals are shorter than 15 minutes. More precisely, in trajectories from the lower-granularity MACACO data, 40% and 65% of GPS points are separated by less than 1 and 5

[3]The rationale behind our choice is that many human activities have been shown to have a weekly periodicity [12], [13]. Using one-week GPS logs lets us capture both repetitiveness and regularity of human mobility.



(a) MACACO (b) OpenStreetMap (c) GeoLife (d) Full view (e) Zoomed view

Fig. 4: (a)-(c) User distribution of one-week trajectories in our reference dataset. (d)-(e) Distribution of the sampling interval in one-week trajectories in our reference dataset: full interval span and sub-minute intervals. Figure best seen in color.

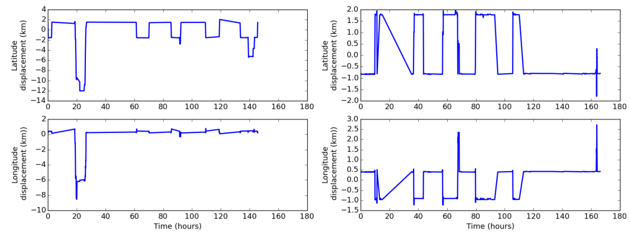


Fig. 5: Unidimensional movement signals of two representative one-week trajectories (left and right). Plots refer to latitude (top) and longitude (bottom) displacements.

minutes, respectively. In trajectories from OSM and GeoLife data, 90% to 95% of points are less than 10 seconds apart, as highlighted in Fig. 4(e). We argue that the sampling intervals in the one-week trajectories of our reference dataset are sufficient to capture human movements, and we consider them as our ground-truth in the remainder of the study.

IV. Spectral analysis of human mobility

From a spectral analysis viewpoint, answering the question posed in the Introduction, *"at what frequency should one periodically sample individual human movements so that they can be reconstructed from the collected samples with minimum loss of information?"*, means to consider human movements as a signal in time, and study its spectrum in frequency.

A. Individual movement as a unidimensional signal

First, we need to transform individual GPS trajectories into unidimensional time series. Even when ignoring altitude information, points in geographical trajectories are obviously bidimensional. We carried out an extensive evaluation of approaches to reduce bidimensional movements to unidimensional signals, using approximated measures such as velocity or relative displacement from the centre of mass, and transformations such as enumeration of discretized locations in the Hilbert space. However, all of the techniques we tested introduced an exceeding amount of noise in the process, disrupting individual movements or introducing unrealistic jumps in the mobility of users.

We opted for a parallel study of the two dimensions of the geographical space, by considering them in isolation. Instead of using the absolute values of latitude and longitude as unidimensional time series, we replace them with the signed latitude and longitude displacements from the corresponding centre of mass of the one-week trajectory. Formally, the displacements of the i-th point in a trajectory are denoted as $\widetilde{lat}_i$ and $\widetilde{lon}_i$, respectively, and computed as

$$\widetilde{lat}_i = lat_i - \frac{1}{n}\sum_{j=0}^{n} lat_j \quad \text{and} \quad \widetilde{lon}_i = lon_i - \frac{1}{n}\sum_{j=0}^{n} lon_j \quad (1)$$

Latitude	Longitude
MACACO	
0.55	0.84
OpenStreetMap	
0.53	0.7
GeoLife	
0.57	0.62

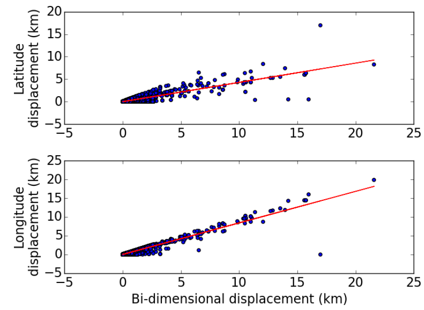


Fig. 6: Correlation between unidimensional displacements and original travelled distance in the bidimensional space. (left) Coefficients separated by data source. (right) Scatterplot of latitude (top) and longitude (bottom) displacemnets with respect to the actual bidimensional movement, for trajectories in the MACACO data.

where lat_j, lon_j are the latitude and longitude coordinates of the j-th GPS point, and n is the number of points in the trajectory. Other than making time series more easily comparable and readable across users and weeks, the transformations in (1) have the property of generating zero-mean signals whose frequency spectra have no DC components. Note that none of the other tested approaches exhibit such properties. Illustrations of our unidimensional description of individual movements are in Fig. 5, for two one-week trajectories.

By considering the transformation above on the two geographical dimensions in isolation we do not introduce errors; yet, we may lose properties that only emerge when the two dimensions are considered jointly. To verify whether such a problem exists, we analysed the correlation between the isolated latitude or longitude displacements and the actual travelled distance in the bidimensional space. Fig. 6 shows the per-source correlation coefficients, as well as the linear fitting on trajectories from the MACACO data. We observe consistently good correlations in all cases, which lets us conclude that both dimensions, when taken separately, still provide decent approximations of the overall mobility. Interestingly, the correlation is always stronger for longitude than for latitude, indicating that participants to all data sources tend to move along an East-West axis rather than along a South-North one; this may be due to the geographical shape of the cities where the data collection took place.

B. Frequency spectra of human mobility

We apply the Fast Fourier Transform (FFT) in order to compute the spectral representation of the finite-length sequences that represent the one-week latitude and longitude

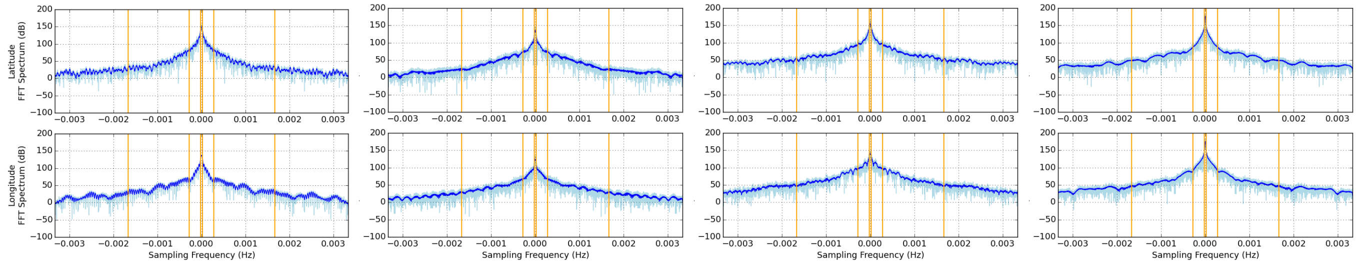


Fig. 7: Frequency spectra of unidimensional movement signals, for a selection of one-week trajectories. Figure best seen in color.

displacement signals. Let $x[n]$, $n = 0, \dots, N-1$ be a sequence (or signal) of length N. Then, its spectral counterpart,

$$X[k] = \sum_{n=0}^{N-1} x[n]\mathrm{e}^{-j2\pi kn/N}, \quad k = 0, \dots, N-1 \qquad (2)$$

can be computed with $\mathcal{O}(N \log N)$ operations. It is well known that $X[k]$ is useful to highlight periodic components in the sequence $x[n]$. In particular, if the sequence $x[n]$ is obtained by sampling a continuous-time periodic signal $x(t)$ within a single period, then $X[k]$ will be equal to the sequence of coefficients of the Fourier series of $x(t)$.

The frequency spectrum of a signal yields information about the sampling frequency needed to reconstruct the original time series with a small error. For an ideal signal, whose spectrum drops to zero after some frequency threshold f_s (i.e., the bandwidth of the signal), the Nyquist–Shannon sampling theorem guarantees that a sampling rate $2f_s$ is enough to allow a lossless reconstruction of the original signal from its samples. For practical signals, the spectrum is not strictly limited, but it features limited amounts of noise. In those cases, the spectrum is mostly concentrated within a finite support and shows a negligible amount of power beyond the frequency threshold; again, sampling at a rate twice the threshold, allows reconstructing the original signal with minimum error.

Fig. 7 shows the spectra of the latitude (top) and longitude (bottom) displacement signals of a representative selection of one-week trajectories. The first two columns refer to the signals in Fig. 5. The original spectra are in light blue, while a moving-average that better displays the overall trends is in dark blue. Vertical orange lines outline the frequencies that correspond to sampling intervals of 10 minutes (farthest from the central frequency), 1 hour and 12 hours (closest to the central frequency). We make two important remarks: *(i)* despite the diversity of the latitude and longitude displacement time series across the different one-week trajectories, all spectra have very similar shapes; *(ii)* the shapes do not show evidence of a bandwidth threshold beyond which the spectra become clearly negligible, making it impossible to identify an operational point for effective sampling.

Due to space limitations, we cannot show the spectra of all one-week trajectories in our dataset: however, we found the observations above to hold in the overwhelming majority of cases. Considering the heterogeneity of our user base, we hypothesize that such features could be a universal property of human mobility spectra.

We can explain both facts remarked above by considering that the unidimensional movement signals show very steep transitions and deep spikes, see, e.g., Fig. 5: hence, the resulting spectrum shows a slow decay for high frequencies. In other words, although it exhibits some clear periodicity [14], human mobility is in fact a sequence of long periods where individuals are almost static and fast transitions between such important locations. While positions during stationary time intervals contribute to low-frequency spectral components and are hence easily captured by a sparse sampling, travelling causes discontinuities in the mobility signal and is much harder to sample. As a result, considering that sampling at higher frequencies has a cost but leads to a better quality of the reconstructed signal, the spectra do not reveal whether, e.g., collecting samples at every 10 minutes is obviously more efficient than sampling at every hour.

Although disappointing in a sense, this outcome calls forth for an extensive quantitative analysis of the exact trade-off between the quality and cost of sampling in the context of human mobility. We address this aspect in the next section.

V. A QUANTITATIVE ANALYSIS OF MOBILITY SAMPLING

We perform an experimental analysis, and investigate the impact of different sampling frequencies on the quality of the mobility reconstructed from the collected samples. To this end, we create downsampled versions of the one-week trajectories in our reference dataset, using a wide range of sampling intervals, from 10 minutes to 12 hours. We deem longer intervals unreasonable for one-week-long time series. We then create reconstructed versions of the complete trajectories by linearly interpolating the samples, and assess how such reconstructed trajectories compare to the original ones.

We measure the error in retrieving a complete individual trajectory from sampled data by using the average Haversine distance. Given two points on Earth's surface, $p_a = (lat_a, lon_a)$ and $p_b = (lat_b, lon_b)$, we define $\Delta_{lat} = lat_b - lat_a$, and $\Delta_{lon} = lon_b - lon_a$. The Haversine distance of p_a and b_b is

$$D(p_a, p_b) = R \cdot 2 \cdot atan2\left(\sqrt{\phi}, \sqrt{(1-\phi)}\right), \qquad (3)$$

with $\phi = \sin^2(\Delta_{lat}/2) + \cos(lon_a) \cdot \cos(lat_b) \cdot \sin^2(\Delta_{lon}/2)$, and $R = 6,371km$ is the Earth's radius. The average Haversine distance of a one-week trajectory is the mean of all Haversine distances between the points of the reconstructed and original mobility recorded at the same time instant.

Fig. 8 shows the evolution of the average Haversine error against the sampling interval, for a representative set of eight individuals in our reference dataset. Each plot presents results for all of the one-week trajectories of a specific user: as multiple one-week trajectories are aggregated in every plot, we outline the mean (dots), 25-75% quantiles (dark blue region) and 10-90% quantiles (light blue region) of the error measured over all trajectories of one user.

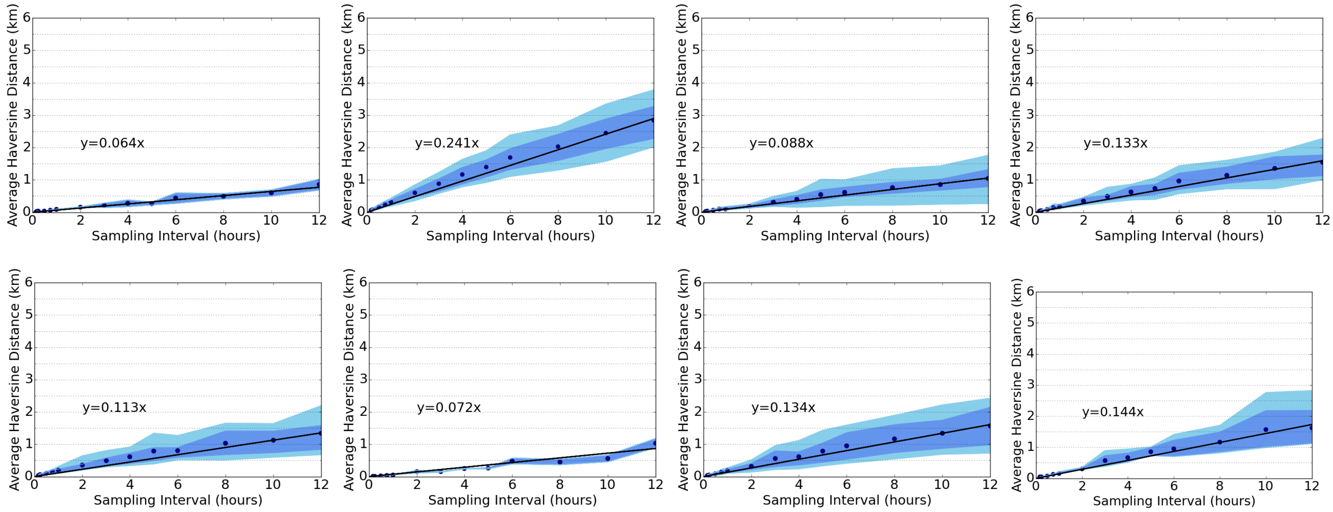
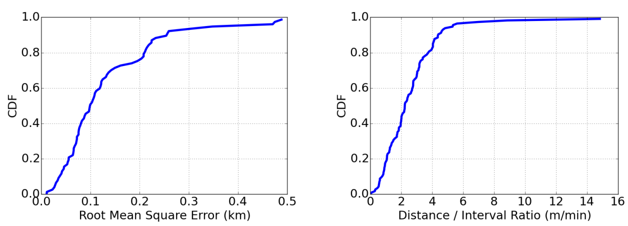


Fig. 8: Average Haversine distance between the original and the reconstructed trajectories of eight users, versus sampling intervals between 10 minutes and 12 hours. Dots represent mean values. Dark and light shaded regions depict the 25-75% and 10-90% quantiles. Solid lines are the linear fittings on average points. Figure best seen in color.



(a) Quality of fitting (b) Fitting slope across users

Fig. 9: (a) Distribution of the RMSE due to the linear approximation of the relation between the average Haversine distance and the sampling interval. (b) Distribution of the ratio between the average Haversine distance and the sampling interval. Plots are for all users.

A surprisingly clear linear relationship characterizes all curves. Fittings on a simple linear model (solid lines in Fig. 8) show an excellent match for all our users. In fact, the linearity of the relationship between the Haversine distance and the sampling interval is a common trait of all individuals in our dataset. Fig. 9(a) portrays the CDF of the Root Mean Square Error (RMSE) between the linear fitting and the mean values (solid lines and dots respectively in Fig. 8) of the average Haversine distance, for all users, over sampling intervals that range from 10 minutes to 12 hours. The probability mass of the distribution is below 250 meters – a very reasonable RMSE for people travelling tens of km per day.

We also highlight that the only parameter of the fitting curve $y = \alpha x$, i.e., the slope α has an important physical meaning: it characterizes the ratio between the average Haversine distance and the sampling interval, or, equivalently, it explains the mean additional error of the reconstructed trajectory when increasing the time that intercurs between samples. Hence, it can be measured in meters per minute (m/min). In other words, our analysis indicates that *adding one minute to the sampling interval used to track one individual leads to an additional positioning error of α minutes in her recorded trajectory, not matter the absolute span of the sampling interval.*

When looking at the value of α, we remark that it is not identical across users: the plots in Fig. 8 also report the equation of the linear fit, and we can note some diversity there. We study the heterogeneity of α in Fig. 9(b), which portrays the CDF of the distance/interval ratio associated to all individuals in our reference dataset. Over 90% of users have slopes that are uniformly distributed between 1 and 4 m/min. Hence, *for the vast majority of individuals, the inaccuracy of their recorded trajectory grows of 1 to 4 meters for each minute added to their movement sampling interval.*

To better understand the reasons behind the linear relationship and the heterogeneity of the α parameter, we computed the correlation between a few human mobility features, such as radius of gyration, average speed, maximum displacement, regularity etc. and the parameter α for all users in our datasets. We found the strongest positive correlation to exist between the radius of gyration and the parameter α, suggesting that the furthest a user is travelling in her mobility the largest the average error she imposes for sparser sampling intervals.

VI. Discussion and Conclusions

Summarizing our findings, we assert that *the average error incurred by trajectories reconstructed from periodic samples scales linearly with the constant sampling interval*. Note that this result is well aligned with the outcome of our spectral analysis in Sec. IV: the linearity of the relationship between error and sampling interval explains the absence of an operational point for the effective sampling of human movements. In addition, we find that *the linear scaling law is characterized by a comparable parameter, i.e., the error-to-interval ratio, across all our user base*; depending on the individual, the error typically grows 1 to 4 meters when adding one minute to the inter-sample time. All this results in a simple and seemingly general scaling law which provides a very useful reference for practitioners: we discuss a few especially promising usages.

Energy-efficient mobile computing. It has been shown that frequent sampling of GPS data tends to quickly drain the battery of a mobile device [15], [16]. A natural solution is to sample the device position at a reduced frequency. However, deciding which periodicity should be employed is not trivial. The linear scaling law we identified is a starting point to understand how to control the tradeoff between energy consumption and localization accuracy. It can also become an important building block in techniques for the dynamic adaptation of GPS data collection frequency in mobile devices –e.g., based not only on context, but also on target accuracy.

Location-based service operation. Location-based services (LBS) rely on positioning information about their users. Yet, an excessively frequent collection of user locations is expensive from both energy and communication perspectives, it raises privacy concerns, and it can ultimately bother customers. Our results may help taking informed decisions, reducing the periodicity of localization according to the approximation in the user's position that can be tolerated by the service.

Active probing of device position in mobile networks. Precise knowledge of subscribers' locations is a valuable information for mobile operators, for both network management and value-added service development [17]. Actively probing mobile devices for their position in a country-scale mobile network is a computationally expensive task, which has traditionally pushed operators to favor less controllable passive measurements [18]. In fact, subscribers are localized based on their associated antenna sector, and those sectors cover hundreds of m^2 in the best case. In this context, our results suggest that running active probing at, e.g., every hour, would not decrease significantly the measurement accuracy, as the incurred error would be typically comparable to the antenna sector coverage.

Trajectory data compression. A straightforward application of our results is data compression. If large amounts of trajectories must be stored, and memory becomes an issue, one could sample the original movement data at some reduced fixed frequency, and store only those samples. This is a lossy operation, yet the scaling law we identified provides reasonable indications about the incurred loss of information.

As a final point of our discussion, we would like to clarify the scope and limitations of our work. The question we posed in the Introduction and the subsequent analyses aim at being as general as possible. However, when it comes to specific applications that leverage some form of human mobility sampling, most have peculiarities that make them unique. For instance, different services may rely on forms of accuracy that are not well represented by our average Haversine error; they may have requirements in terms of maximum variance of the accuracy that is not captured by our mean analysis; or they may accomodate sampling approaches that are not constant but adaptive to, e.g., context. It is thus important to understand that *our study does not aim at providing direct support to the design of specific practical services*: instead, our scaling law can be a sensible starting point for more application-targeted investigations of trajectory sampling.

In the light of the results presented in this work, a laconic answer to the original question posed in the Introduction is *"it depends on the error one can afford"*. Fortunately, our findings are more informative than that, and provide a simple scaling law for the user positioning error, with general validity and limited parameter diversity across individuals.

REFERENCES

[1] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi, "Understanding individual human mobility patterns," *Nature*, vol. 453, no. 7196, pp. 779–782, June 2008.

[2] P. Katsikouli, R. Sarkar, and J. Gao, "Persistence based online signal and trajectory simplification for mobile devices," in *Proceedings of the 22Nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, ACM, 2014, pp. 371–380.

[3] J. Muckell, J.-H. Hwang, C. T. Lawson, and S. S. Ravi, "Algorithms for compressing gps trajectory data: An empirical evaluation," in *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, ACM, 2010, pp. 402–405.

[4] B. Thierry, B. Chaix, and Y. Kestens, "Detecting activity locations from raw gps data: a novel kernel-based algorithm," *International Journal of Health Geographics*, vol. 12, no. 1, p. 14, 2013.

[5] B. C. Csáji, A. Browet, V. Traag, J.-C. Delvenne, E. Huens, P. V. Dooren, Z. Smoreda, and V. D. Blondel, "Exploring the mobility of mobile phone users," *Physica A: Statistical Mechanics and its Applications*, vol. 392, no. 6, pp. 1459 – 1473, 2013.

[6] Y. Chen, K. Jiang, Y. Zheng, C. Li, and N. Yu, "Trajectory simplification method for location-based social networking services," in *Proceedings of the 2009 International Workshop on Location Based Social Networks (LBSN)*, ACM, 2009, pp. 33–40.

[7] S. Khetarpaul, R. Chauhan, S. K. Gupta, L. V. Subramaniam, and U. Nambiar, "Mining gps data to determine interesting locations," in *Proceedings of the 8th International Workshop on Information Integration on the Web: In Conjunction with WWW 2011 (IIWeb)*, ACM, 2011, pp. 8:1–8:6.

[8] C. Long, R. C.-W. Wong, and H. V. Jagadish, "Direction-preserving trajectory simplification," *Proc. VLDB Endow.*, vol. 6, no. 10, pp. 949–960, Aug. 2013.

[9] G. Trajcevski, H. Cao, P. Scheuermanny, O. Wolfsonz, and D. Vaccaro, "On-line data reduction and the quality of history in moving objects databases," in *Proceedings of the 5th ACM International Workshop on Data Engineering for Wireless and Mobile Access (MobiDE)*, ACM, 2006, pp. 19–26.

[10] R. Lange, F. Dürr, and K. Rothermel, "Online trajectory data reduction using connection-preserving dead reckoning," in *Proceedings of the 5th Annual International Conference on Mobile and Ubiquitous Systems: Computing, Networking, and Services (Mobiquitous)*, ICST, 2008, pp. 52:1–52:10.

[11] Y. Zheng, Q. Li, Y. Chen, X. Xie, and W.-Y. Ma, "Understanding mobility based on gps data," in *Proceedings of the 10th International Conference on Ubiquitous Computing (UbiComp)*, ACM, 2008, pp. 312–321.

[12] R. Keralapura, A. Nucci, Z.-L. Zhang, and L. Gao, "Profiling users in a 3g network using hourglass co-clustering," in *Proceedings of the Sixteenth Annual International Conference on Mobile Computing and Networking (MobiCom)*, ACM, 2010, pp. 341–352.

[13] M. Z. Shafiq, L. Ji, A. X. Liu, and J. Wang, "Characterizing and modeling internet traffic dynamics of cellular devices," in *Proceedings of the ACM SIGMETRICS Joint International Conference on Measurement and Modeling of Computer Systems*, ACM, 2011, pp. 305–316.

[14] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási, "Limits of predictability in human mobility," *Science*, vol. 327, no. 5968, pp. 1018–1021, 2010.

[15] Y. Qi, C. Yu, Y. J. Suh, and S. Y. Jang, "Gps tethering for energy conservation1," in *2015 IEEE Wireless Communications and Networking Conference (WCNC)*, March 2015, pp. 1320–1325.

[16] S. Vanini, F. Faraci, A. Ferrari, and S. Giordano, "Using barometric pressure data to recognize vertical displacement activities on smartphones," *Computer Communications*, vol. 87, pp. 37 – 48, 2016.

[17] A. Finamore, M. Mellia, Z. Gilani, K. Papagiannaki, V. Erramilli, and Y. Grunenberger, "Is there a case for mobile phone content pre-staging?" in *Proceedings of the Ninth ACM Conference on Emerging Networking Experiments and Technologies (CoNEXT)*, ACM, 2013, pp. 321–326.

[18] M. Ficek, T. Pop, P. Vláčil, K. Dufková, L. Kencl, and M. Tomek, "Performance study of active tracking in a cellular network using a modular signaling platform," in *Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services (MobiSys)*, ACM, 2010, pp. 239–254.